RETROVIRAL NUCLEIC MATERIAL AND NUCLEOTIDE FRAGMENTS, IN PARTICULAR ASSOCIATED WITH MULTIPLE SCLEROSIS AND/OR RHEUMATOID ARTHRITIS, FOR DIAGNOSTIC, PROPHYLACTIC AND THERAPEUTIC USES

Multiple sclerosis (MS) is a demyelinizing disease of the central nervous system (CNS) of which the complete cause still remains unknown.

Numerous studies have supported the hypothesis for a viral etiology of the disease, but none of the known viruses tested has proved to be the causative agent tested for: a review of the viruses tested for in MS for many years has been carried out by E. Norrby and R.T. Johnson.

Recently, a retrovirus, different from the known human retroviruses, was isolated from patients suffering from MS. The authors were able to show that this retrovirus could be transmitted in vitro, that patients suffering from MS produced antibodies capable of recognizing proteins associated with the infection of the leptomeningeal cells by this retrovirus, and that the expression of the latter could be greatly stimulated by the immediate-early genes of some herpesviruses.

All these results argue in favor of the role in MS of at least one unknown retrovirus or of a virus having a reverse transcriptase (RT) activity which is detectable by the method published by H. Perron and termed "LM7-type RT" activity.

The studies by the applicant have made it possible to obtain two continuous cell lines infected with natural isolates obtained from two different patients suffering from MS, by a culture method as described in the document WO-A-93 20188, whose content is incorporated by reference into the present description. These two lines derived from cells of human choroid plexus, called LM7PC and PLI-2, were deposited at the E.C.A.C.C. on 22 July 1992 and 8 January 1993, respectively, under numbers 92 072201

and 93 010817, in accordance with the provisions of the Treaty of Budapest. Moreover, the viral isolates possessing an LM7-type RT activity have also been deposited at the E.C.A.C.C. under the overall name of "strains". The "strain" or isolate harbored by the PLI-2 line, called POL-2, was deposited at the E.C.A.C.C. on 22 July 1992 under No. V92072202. The "strain" or isolate harbored by the LM7PC line, called MS7PG, was deposited at the E.C.A.C.C. on 8 January 1993 under No. V93010816.

Using the abovementioned cultures and isolates, characterized by biological and morphological criteria, efforts were then made to characterize the genetic material associated with the viral particles produced in these cultures.

The proportions of genome already characterized were used to develop molecular detection tests for the viral genome and immunoserological tests, using the amino acid sequences encoded by the nucleotide sequences of the viral genome, in order to detect the immune response directed against epitopes associated with the viral infection and/or expression.

These tools have already made it possible to confirm an association between MS and the expression of the sequences identified in the patents cited further on. However, the viral system discovered by the applicant is related to a complex retroviral system. Indeed, the sequences which are found to be encapsidated in the extracellular viral particles produced by the different cultures of cells of patients suffering from MS show clearly that there is co-encapsidation of retroviral genomes which are related but different from the "wild-type" retroviral genome which produces the infectious viral particles. This phenomenon was observed between replicative retroviruses and endogenous retroviruses belonging to the same family, or even heterologous retroviruses. The concept of endogenous retrovirus is very important in the context of our discovery because, in the case of

MSRV-1, it has been observed that endogenous retroviral sequences comprising sequences homologous to the MSRV-1 genome exist in normal human DNA. The existence of endogenous retroviral elements (ERV) related to MSRV-1 through all or part of their genome explains the fact that the expression of the MSRV-1 retrovirus in human cells can interact with related endogenous sequences. These interactions are found in the case of pathogenic and/or infectious endogenous retroviruses (for example some ecotropic strains of the Murine Leukemia virus), in the case of exogenous retroviruses whose nucleotide sequence may be found partially or completely in the form of ERVs, in the genome of the host animal (e.g. mouse mammary tumor exogenous virus transmitted via milk). These interactions consist mainly of (i) a transactivation or co-activation of ERVs by the replicative retrovirus, (ii) an "illegitimate" encapsidation of related RNAs of ERVs, or of ERVs - or even of cellular RNAs - simply possessing compatible encapsidation sequences, into the retroviral particles produced by the expression of the replicative strain, which are sometimes transmissible and sometimes with an inherent pathogenicity, and (iii) relatively high recombinations between the co-encapsidated genomes, in particular in the reverse transcription phases, which lead to the formation of hybrid genomes, which are sometimes transmissible and sometimes with an inherent pathogenicity.

Thus, (i) various MSRV-1-related sequences have been found in purified viral particles; (ii) molecular analysis of the various regions of the MSRV-1 retroviral genome should be carried out by systematically analyzing the co-encapsidated, interfering and/or recombinant sequences which are generated by the infection and/or expression of MSRV-1; furthermore, some clones may have portions of defective sequences produced by the retroviral replication and the template and/or transcription errors caused by reverse transcriptase; (iii) the families of sequences

related to the same retroviral genomic region are the supports for an overall diagnostic detection which may be optimized by the identification of invariable regions among the clones expressed and by the identification of reading frames responsible for the production of antigenic and/or pathogenic polypeptides which may only be produced by a portion, or even only one, of the clones expressed and under these conditions, the systematic analysis of the clones expressed in one region of a given gene makes it possible to evaluate the frequency of variation and/or recombination of the MSRV-1 genome in this region and to define the optimum sequences for the applications, in particular the diagnostic applications; (iv) the pathology caused by a retrovirus such as MRSV-1 may be a direct effect of its expression and of the proteins or peptides produced as a result, but also an effect of the activation, encapsidation, recombination of related or heterologous genomes and proteins or peptides produced as a result; thus, these genomes associated with the expression and/or infection by MSRV-1 are an integral part of the potential pathogenicity of this virus and therefore constitute diagnostic detection supports and particular therapeutic targets. Likewise, any agent which is associated with, or which is a cofactor for these interactions responsible for the pathogenicity in question, such as MSRV-2 or the gliotoxic factor described in the patent application published under the No. FR-2,716,198, can participate in the development of an overall and very effective strategy for therapeutic diagnosis, prognosis, monitoring and/or integrated therapy for MS in particular, but also for any other disease associated with the same agents.

In this context, a parallel discovery has been made in another autoimmune disease, rheumatoid arthritis (RA), which has been described in the French patent application filed under the No. 95 02960. This discovery shows that, by applying methodological

approaches similar to those which were used in the studies by the applicant on MS, it has been possible to identify a retrovirus expressed in RA which shares the sequences described for MSRV-1 in MS and also the co-existence of an MSRV-2-associated sequence which is also described in MS. As regards MSRV-1, the sequences commonly detected in MS and RA relate to the *pol* and *gag* genes. On the basis of current knowledge, it is possible to combine the *gag* and *pol* sequences described with the MSRV-1 strains expressed in these two diseases.

The present patent application has as its object various results, supplementary in relation to those already protected by the French patent applications:

- No. 92/04322 of 03.04.1992, published under No. 2,689,519;
- No. 92/13447 of 03.11.1992, published under No. 2,689,521;
- No. 92/13443 of 03.11.1992, published under No. 2,689,520;
- No. 94/01529 of 04.02.1994, published under No. 2,715,936;
- No. 94/01531 of 04.02.1994, published under No. 2,715,939;
- No. 94/01530 of 04.02.1994, published under No. 2,715,936;
- No. 94/01532 of 04.02.1994, published under No. 2,715,937;
- No. 94/14322 of 24.11.1994, published under No. 2,727,428;
- No. 94/15810 of 23.12.1994, published under No. 2,728,585;

and

- Patent Application WO 97/06260.

The present invention relates, first of all, to a nucleic material, which may consist of a retroviral material, in isolated or purified state, which may be understood or characterized in various ways:

- it comprises a nucleotide sequence chosen from the group which consists of (i) the sequences SEQ ID NO: 112, SEQ ID NO: 114, SEQ ID NO: 117, SEQ ID NO: 120, SEQ ID NO: 124, SEQ ID NO: 130, SEQ ID NO: 141 and SEQ ID NO: 142; (ii) the sequences complementary to sequences (i); and (iii) the sequences equivalent to sequences (i) or (ii), in particular the sequences having, for every series of 100 contiguous monomers, at least 50%, and preferentially at least 70% homology with sequences (i) or (ii) respectively;

- it encodes a polypeptide having, for every contiguous series of at least 30 amino acids, at least 50%, and preferably at least 70% homology with a peptide sequence chosen from the group which consists of SEQ ID NO: 113, SEQ ID NO: 115, SEQ ID NO: 118, SEQ ID NO: 121, SEQ ID NO: 135 and SEQ ID NO: 137;

- its pol gene comprises a nucleotide sequence identical or equivalent to a sequence chosen from the group which consists of SEQ ID NO: 112, SEQ ID NO: 124 and their complementary sequences;

- the 5' end of its pol gene starts at nucleotide 1419 of SEQ ID NO: 130;

- its pol gene encodes a polypeptide having, for every contiguous series of at least 30 amino acids, at least 50%, and preferably at least 70% homology with the peptide sequence SEQ ID NO: 113;

- the 3' end of its gag gene ends at nucleotide 1418 of SEQ ID NO: 130;

- its env gene comprises a nucleotide sequence identical or equivalent to a sequence chosen from the group which consists of SEQ ID NO: 117, and its complementary sequences;

- its env gene comprises a nucleotide sequence which starts at nucleotide 1 of SEQ ID NO: 117 and ends at nucleotide at nucleotide [sic] 233 of SEQ ID NO: 114;

- its env gene encodes a polypeptide having, for every contiguous series of at least 30 amino acids,

at least 50%, and preferably at least 70% homology with the sequence SEQ ID NO: 118;

- the U3R region of its 3' LTR comprises a nucleotide sequence which ends at nucleotide 617 of SEQ ID NO: 114;

- the RU5 region of its 5' LTR comprises a nucleotide sequence which starts at nucleotide 755 of SEQ ID NO: 120 and ends at nucleotide 337 of SEQ ID NO: 141 or SEQ ID NO: 142;

- a retroviral nucleic material comprising a sequence which starts at nucleotide 755 of SEQ ID NO: 120 and which ends at nucleotide 617 of SEQ ID NO: 114;

- the retroviral nucleic material as defined above is in particular associated with at least one autoimmune disease such as multiple sclerosis or rheumatoid arthritis.

The invention also relates to a nucleotide fragment which corresponds to at least one of the following definitions:

- it comprises or consists of a nucleotide sequence chosen from the group which consists of (i) the sequences SEQ ID NO: 112, SEQ ID NO: 114, SEQ ID NO: 117, SEQ ID NO: 120, SEQ ID NO: 124, SEQ ID NO: 130, SEQ ID NO: 141 and SEQ ID NO: 142; (ii) the sequences complementary to sequences (i); and (iii) the sequences equivalent to sequences (i) or (ii), in particular the sequences having, for every series of 100 contiguous monomers, at least 50%, and preferentially at least 70% homology with sequences (i) or (ii) respectively;

- it comprises or consists of a nucleotide sequence encoding a polypeptide having, for every contiguous series of at least 30 amino acids, at least 50%, and preferably at least 70% homology with a peptide sequence chosen from the group which consists of SEQ ID NO: 113, SEQ ID NO: 115, SEQ ID NO: 118, SEQ ID NO: 121, SEQ ID NO: 135 and SEQ ID NO: 137.

Other subjects of the present invention are the following:

- a nucleic probe for the detection of a retrovirus associated with multiple sclerosis and/or rheumatoid arthritis, capable of hybridizing specifically with any fragment defined above and belonging to the genome of said retrovirus; it advantageously possesses from 10 to 100 nucleotides, preferably from 10 to 30 nucleotides;

- a primer for the amplification, by polymerization, of an RNA or of a DNA of a retrovirus associated with multiple sclerosis and/or rheumatoid arthritis, which comprises a nucleotide sequence identical or equivalent to at least a portion of the nucleotide sequence of a fragment defined above, in particular a nucleotide sequence having, for every series of 10 contiguous monomers, at least 50%, preferably at least 70% homology with at least said portion of said fragment; preferably the nucleotide sequence of a primer of the invention is chosen from SEQ ID NO: 116, SEQ ID NO: 119, SEQ ID NO: 122, SEQ ID NO: 123, SEQ ID NO: 126, SEQ ID NO: 127, SEQ ID NO: 128, SEQ ID NO: 129, SEQ ID NO: 132, and SEQ ID NO: 133;

- an RNA or a DNA, and in particular a replication and/or expression vector, comprising a genomic fragment of the nucleic material or a fragment defined above;

- a peptide encoded by any open reading frame belonging to a nucleotide fragment defined above, in particular a polypeptide, for example oligopeptide forming or comprising an antigenic determinant recognized by sera of patients infected with the MSRV-1 virus, and/or in whom the MSRV-1 virus has been reactivated; a preferential peptide comprises a sequence identical, partially or completely, or equivalent to a sequence chosen from SEQ ID NO: 113, SEQ ID NO: 115, SEQ ID NO: 118, SEQ ID NO: 121, SEQ ID NO: 135 and SEQ ID NO: 137;

- a diagnostic, prophylactic or therapeutic composition, in particular for inhibiting the

expression of at least one retrovirus associated with multiple sclerosis and/or rheumatoid arthritis, comprising a nucleotide fragment defined above;

- a method for detecting a retrovirus associated with multiple sclerosis and/or rheumatoid arthritis, in a biological sample, comprising the steps consisting of bringing an RNA and/or a DNA assumed to belong to or obtained from said retrovirus, or their complementary RNA and/or DNA, into contact with a composition comprising a nucleotide fragment defined above.

Before detailing the invention, various terms used in the description and the claims are now defined:

- strain or isolate is understood to mean any infectious and/or pathogenic biological fraction containing, for example, viruses and/or bacteria and/or parasites, generating a pathogenic and/or antigenic power, harbored by a culture or a live host; by way of example, a viral strain according to the preceding definition may contain a co-infectious agent, for example a pathogenic protist,

- the term "MSRV" used in the present description designates any pathogenic and/or infectious agent, as associated with MS, in particular a viral species, the attenuated strains of said viral species, or the interfering defective particles or particles containing co-encapsidated genomes or alternatively genomes recombined with a portion of the MSRV-1 genome, which are derived from this species. It is known that viruses and particularly viruses containing RNA exhibit variability, following in particular relatively high rates of spontaneous mutation, which will be taken into account below to define the concept of equivalence,

- human virus is understood to mean a virus capable of infecting or of being harbored by human beings,

- given all the natural or induced variations and/or recombination which may be encountered in practice in the present invention, the objects thereof,

defined above and in the claims, have been expressed by comprising the equivalents or derivatives of the various biological materials defined below, in particular homologous nucleotide or peptide sequences,

- the variant of a virus or of a pathogenic and/or infectious agent according to the invention comprises at least one antigen recognized by at least one antibody directed against at least one corresponding antigen of said virus and/or of said pathogenic and/or infectious agent, and/or a genome in which any portion is detected by at least one hybridization probe, and/or at least one nucleotide amplification primer specific for said virus and/or pathogenic and/or infectious agent, under defined hybridization conditions well known to persons skilled in the art,

- according to the invention, a nucleotide fragment or an oligonucleotide or a polynucleotide is a stretch of monomers, or a biopolymer, characterized by the informational sequence of the natural nucleic acids, which is capable of hybridizing to any other nucleotide fragment under predefined conditions, it being possible for the stretch to contain monomers of different chemical structures and to be obtained from a natural nucleic acid molecule and/or by genetic recombination and/or by chemical synthesis; a nucleotide fragment may be identical to a genomic fragment of the MSRV-1 virus considered by the present invention, in particular a gene of the latter, for example pol or env in the case of said virus;

- thus, a monomer may be a natural nucleic acid nucleotide in which the constituent components are a sugar, a phosphate group and a nitrogen base; in RNA, the sugar is ribose; in DNA, the sugar is 2-deoxy-ribose; depending on whether DNA or RNA is involved, the nitrogen base is chosen from adenine, guanine, uracil, cytosine, thymine; or the nucleotide may be modified in at least one of the three constituent components; by way of example, the modification may

occur at the level of the bases, generating modified bases such as inosine, 5-methyl-deoxycytidine, deoxyuridine, 5-dimethylamineodeoxyuridine [sic], 2,6-diamineopurine [sic], 5-bromodeoxyuridine and any other modified base promoting hybridization; at the level of the sugar, the modification may consist in the replacement of at least one deoxyribose with a polyamide, and at the level of the phosphate group, the modification may consist in its replacement with esters, in particular chosen from the esters of diphosphate, of alkyl and arylphosphonate and of phosphorothioate,

- "informational sequence" is understood to mean any ordered series of monomers, whose chemical nature and in which the order in a reference direction, constitute or otherwise a functional information of the same quality as that for the natural nucleic acids,

- hybridization is understood to mean the process during which, under appropriate operating conditions, two nucleotide fragments, having sufficiently complementary sequences, become annealed to form a complex, in particular a double or triple, structure, preferably in helical form,

- a probe comprises a nucleotide fragment synthesized by the chemical route or obtained by digestion or enzymatic cleavage of a longer nucleotide fragment, comprising at least six monomers, advantageously from 10 to 100 monomers, preferably 10 to 30 monomers, and possessing a hybridization specificity under defined conditions; preferably, a probe possessing less than 10 monomers is not used alone, but is used in the presence of other probes which are equally short in length or otherwise; under certain specific conditions, it may be useful to use probes which are greater than 100 monomers in size; a probe may be used in particular for diagnostic purposes, and it may be, for example, capture and/or detection probes,

- the capture probe may be immobilized on a solid support by any appropriate means, that is to say directly or indirectly, for example by covalent bonding or passive adsorption,

- the detection probe may be labeled by means of a marker chosen in particular from radioactive isotopes, enzymes chosen in particular from peroxidase and alkaline phosphatase and those capable of hydrolyzing a chromogenic, fluorigenic or luminescent substrate, chromophoric chemical compounds, chromogenic, fluorigenic or luminescent compounds, analogs of nucleotide bases, and biotin,

- the probes used for diagnostic purposes of the invention may be used in all known hybridization techniques, and in particular the so-called "DOT-BLOT" technique, "SOUTHERN BLOT" technique, "NORTHERN BLOT" technique which is a technique identical to the "SOUTHERN BLOT" technique but which uses RNA as target, the SANDWICH technique; advantagously, the SANDWICH technique is used in the present invention, comprising a specific capture probe and/or a specific detection probe, it being understood that the capture probe and the detection probe must have a nucleotide sequence which is at least partially different,

- any probe according to the present invention may hybridize in vivo or in vitro with the RNA and/or with the DNA, in order to block the replication, in particular translation and/or transcription, phenomena and/or to degrade said DNA and/or RNA,

- a primer is a probe comprising at least six monomers, and advantageously from 10 to 30 monomers, possessing hybridization specificity under defined conditions, for the initiation of an enzymatic polymerization, for example in an amplification technique such as PCR (Polymerase Chain Reaction), in an extension method such as sequencing, in a reverse transcription method and the like,

- two nucleotide or peptide sequences are said to be equivalent or derived with respect to each other,

or with respect to a reference sequence, if functionally the corresponding biopolymers can play substantially the same role, without being identical, in relation to the application or use considered, or in the technique in which they are involved; particularly equivalent are two sequences obtained because of the natural variability, in particular spontaneous mutation, of the species from which they were identified, or induced mutation, as well as two homologous sequences, the homology being defined below,

- "variability" is understood to mean any spontaneous or induced modification of a sequence, in particular by substitution, and/or insertion, and/or deletion of nucleotides and/or of nucleotide fragments, and/or extension and/or shortening of the sequence at least at one of the ends; a nonnatural variability may result from the genetic engineering techniques used, for example from the choice of the degenerate or nondegenerate synthetic primers selected to amplify a nucleic acid; this variability may result in modifications of any starting sequence, considered as a reference, and which may be expressed by a degree of homology with respect to said reference sequence,

- homology characterizes the degree of identity of two compared nucleotide or peptide fragments; it is measured by the percentage identity which is in particular determined by direct comparison of nucleotide or peptide sequences, with respect to reference nucleotide or peptide sequences,

- any nucleotide fragment is said to be equivalent to or derived from a reference fragment if it has a nucleotide sequence equivalent to the sequence of the reference fragment; according to the preceding definition, in particular equivalent to a reference nucleotide fragment are:

(a) any fragment capable of hybridizing, at least partially, with the complementary to the reference fragment,

(b) any fragment whose alignment with the reference fragment leads to the identification of identical contiguous bases, in a greater number than with any other fragment obtained from another taxonomic group,

(c) any fragment resulting or capable of resulting from the natural variability of the species from which it is obtained,

(d) any fragment which may result from genetic engineering techniques applied to the reference fragment,

(e) any fragment, containing at least eight contiguous nucleotides, encoding a peptide homologous or identical to the peptide encoded by the reference fragment,

(f) any fragment different from the reference fragment through insertion, deletion, substitution of at least one monomer, extension, or shortening at least at one of its ends; for example, any fragment corresponding to the reference fragment, flanked at least at one of its ends by a nucleotide sequence not encoding a polypeptide,

- polypeptide is understood to mean in particular any peptide of at least two amino acids, in particular oligopeptide, protein, extracted, separated, or substantially isolated or synthesized, through the involvement of humans, in particular those obtained by chemical synthesis, or through expression in a recombinant organism,

- polypeptide partially encoded by a nucleotide fragment is understood to mean a polypeptide having at least three amino acids encoded by at least nine contiguous monomers included in said nucleotide fragment,

- an amino acid is said to be analogous to another amino acid when their respective physicochemical characteristics, such as polarity, hydrophobicity and/or basicity, and/or acidity, and/or

neutrality, are substantially the same; thus, a leucine is analogous to an isoleucine,

- any polypeptide is said to be equivalent to or derived from a reference polypeptide if the polypeptides compared have substantially the same properties, and in particular the same antigenic, immunological, enzymatic and/or molecular recognition properties; in particular equivalent to a reference polypeptide is:

(a) any polypeptide possessing a sequence in which at least one amino acid has been replaced by an analogous amino acid,

(b) any polypeptide having an equivalent peptide sequence, obtained by natural or induced variation of said reference polypeptide, and/or of the nucleotide fragment encoding said polypeptide,

(c) a mimotope of said reference polypeptide,

(d) any polypeptide from whose sequence one or more amino acids of the L series are replaced by an amino acid of the D series, and vice versa,

(e) any polypeptide into whose sequence a modification of the side chains of the amino acids has been introduced, such as for example an acetylation of the amine-containing functions, a carboxylation of the thiol functions, an esterification of the carboxyl functions,

(f) any polypeptide in whose sequence one or more peptide bonds have been modified, such as for example the carba, retro, inverso, retro-inverso, reduced, and methylene-oxy bonds,

(g) any polypeptide in which at least one antigen is recognized by an antibody directed against a reference polypeptide,

- the percentage identity characterizing the homology between two peptide fragments compared is according to the present invention at least 50% and preferably at least 70%.

Given that a virus possessing a reverse transcriptase enzymatic activity may be genetically

characterized both in RNA and DNA form, both the viral DNA and RNA will be mentioned in order to characterize the sequences relative to a virus possessing such a reverse transcriptase activity, termed MSRV-1 according to the present description.

The expressions of order which are used in the present description and the claims, such as "first nucleotide sequence", are not selected to express a particular order, but to define the invention more clearly.

Detection of a substance or agent is understood below to mean an identification, a quantification or a separation or isolation of said substance or of said agent.

The invention will be understood more clearly on reading the detailed description which follows which is made with reference to the appended figures in which:

Figure 1 represents the general structure of the proviral DNA and the genomic RNA of MSRV-1.

Figure 2 represents the nucleotide sequence of the clone called CL6-5' (SEQ ID NO: 112) and three potential reading frames in amino acids presented under the nucleotide sequence.

Figure 3 represents the nucleotide sequence of the clone called CL6-3' (SEQ ID NO: 114) and three potential reading frames in amino acids presented under the nucleotide sequence.

Figure 4 represents the nucleotide sequence of the clone called C15 (SEQ ID NO: 117) and three potential reading frames in amino acids presented under the nucleotide sequence.

Figure 5 represents the nucleotide sequence of the clone called 5M6 (SEQ ID NO: 120) and three potential reading frames in amino acids presented under the nucleotide sequence.

Figure 6 represents the nucleotide sequence of the clone called CL2 (SEQ ID NO: 130) and three

potential reading frames in amino acids presented under the nucleotide sequence.

Figure 7 represents three potential reading frames in amino acids expressed by pET28C-clone 2 and presented under the nucleotide sequence.

Figure 8 represents three potential reading frames in amino acids expressed by pET21C-clone 2 and presented under the nucleotide sequence.

Figure 9 represents the nucleotide sequence of the clone called LB13 (SEQ ID NO: 141) and three potential reading frames in amino acids presented under the nucleotide sequence.

Figure 10 represents the nucleotide sequence of the clone called LA15 (SEQ ID NO: 142) and three potential reading frames in amino acids presented under the nucleotide sequence.

Figure 11 represents the nucleotide sequence of the clone called LB16 (SEQ ID NO: 124) and three potential reading frames in amino acids presented under the nucleotide sequence.

Figure 12 represents the promoter activity expressed in cpm/4 min of the U3R sequences subcloned from LTRs of different origins into the plasmid PCAT3. PCAT3 means plasmid alone, PCAT-PH74 means plasmid plus endogenous U3R clone expressed in the placenta, PCAT-c16 means plasmid plus U3R clone amplified in the RNA of an MS plasma, PCAT-5M6 means plasmid plus U3R region amplified in the cellular DNA, "no plasmid" means absence of plasmid in the test.

Figure 13 represents the MSRV1 env and 3' LTR sequences. The horizontal arrows indicate the start of the env, U3 and R regions. In the env region, the signal peptide and the potential immunosuppressive region are underlined, the potential glycosilation sites are boxed and the potential cleavage sites are indicated by vertical arrows. In the U3R region: the regulatory element CAAT and the TATA Box are underlined, the "cap" site and the polyadenylation signal are also indicated.

Figure 14 represents the 5' LTR (RU5) region followed by a PBS site (primer binding site) complementary to the Trp tRNA and by a gag gene encoding a protein of about 487 amino acids. The amino acids conserved in the nucleocapsid are underlined twice. The amino acids defining the region of greatest homology in the capsid are in bold and underlined once. The / symbols in the amino acid sequence indicate variations observed depending on the clones and, in the nucleotide sequence, they indicate frame jumps in some clones. The boxed regions correspond to epitopes identified by peptide analysis of the C-terminal region.

Figure 15 represents the integrase region of MSRV1, the nucleotide sequence and the amino acid sequence deduced from the integrase region corresponding to clone 87-23. In Figure 15, // means a frame jump which has been suppressed in order to restore the potential ORF. The letters in underlined bold characters represent the conserved amino acids in the retroviral integrases.

Figure 16 describes the nucleotide and peptide sequences of clone B13 (identical to clone FBd13 described in previous applications) with indication of the ORFs and stop codons represented by a dot. The underlined region in bold represents the potential immunosuppressive domain. The single underlined domain represents the start of the 3' LTR.

EXAMPLE 1: PREPARATION OF A CL6-5' REGION ENCODING THE N-TERMINAL END OF INTEGRASE AND OF A CL6-3' REGION CONTAINING THE 3' TERMINAL SEQUENCE OF THE MSRV-1 GENOME

A 3' RACE was carried out on the total RNA extracted from plasma from a patient suffering from MS. A healthy control plasma, treated under the same conditions, was used as negative control. The synthesis of cDNA was carried out with an oligo dT primer identified by SEQ ID NO: 68 (5' GAC TCG CTG CAG ATC GAT

TTT TTT TTT TTT TTT T 3') and the reverse transcriptase "Expand™ RT" from Boehringer according to the conditions recommended by the company. A PCR was carried out with the enzyme Klentaq (Clontech) under the following conditions: 94°C 5 min then 93°C 1 min, 58°C 1 min, 68°C 3 min over 40 cycles and 68°C for 8 min, with a final reaction volume of 50 µl.

Primers used for the PCR:

- 5' primer, identified by SEQ ID NO: 69

5' GCC ATC AAG CCA CCC AAG AAC TCT TAA CTT 3';

- 3' primer, identified by SEQ ID NO: 68

A second so-called "seminested" PCR was carried out with a 5' primer situated inside the region already amplified. This second PCR was carried out under the same experimental conditions as those used for the first PCR, using 10 µl of the amplification product derived from the first PCR.

Primers used for the seminested PCR:

- 5' primer, identified by SEQ ID NO: 70

5' CCA ATA GCC AGA CCA TTA TAT ACA CTA ATT 3';

- 3' primer, identified by SEQ ID NO: 68

The primers SEQ ID NO: 69 and SEQ ID NO: 70 are specific for the pol region of MRSV-1.

An amplification product of 1.9 Kb was obtained for the plasma of the MS patient. The corresponding fragment was not observed for the healthy control plasma. This amplification product was cloned in the following manner:

The amplified DNA was inserted into a plasmid with the aid of the TA Cloning kit®. The 2 µl of DNA solution were mixed with 5 µl of sterile distilled water, 1 µl of a 10 times concentrated ligation buffer "10X LIGATION BUFFER", 2 µl of "pCR™ VECTOR" (25 ng/ml) and 1 µl of "T4 DNA LIGASE". This mixture was incubated overnight at 12°C. The next steps were carried out in accordance with the instructions for the TA Cloning kit® (Invitrogen). At the end of the procedure, the white colonies of recombinant bacteria (white) were subcultured so as to be cultured and allow the

extraction of the plasmids incorporated according to the so-called "miniprep" procedure. The plasmid preparation of each recombinant colony was cut with an appropriate restriction enzyme and analyzed on agarose gel. The plasmids possessing an insert detected under UV light after staining the gel with ethidium bromide were selected for the sequencing of the insert after hybridization with a primer complementary to the Sp6 promoter present on the cloning plasmid of the TA cloning kit®. The reaction prior to the sequencing was then carried out according to the method recommended for using the sequencing kit "PRISM™ Ready Reaction AmpliTaq® FS, DyeDeoxy™ Terminator" (Applied Biosystems, ref. 402119) and the automated sequencing was carried out on the Applied Biosystems 373 A and 377 apparatus, according to the manufacturer's instructions.

The clone obtained contains a CL6-5' region encoding the N-terminal end of integrase and a CL6-3' region corresponding to the 3' terminal region of MSRV-1 and making it possible to define the end of the envelope (234 bp) and the U3 and R (401 bp) regions of the MSRV1 retrovirus.

The region corresponding to the N-terminal end of integrase is represented by its nucleotide sequence (SEQ ID NO: 112) in Figure 27. The three potential reading frames are presented by their amineo [sic] acid sequence under the nucleotide sequence, and the amineo [sic] acid sequence of the N-terminal end of integrase is identified by SEQID NO: 113.

The C16-3' region is represented by its nucleotide sequence (SEQ ID NO: 114) in Figure 3. The three potential reading frames are presented by their amineo [sic] acid sequence under the nucleotide sequence. An amineo [sic] acid sequence corresponding to the C-terminal end of the MSRV-1 env protein is identified by SEQ ID NO: 115.

In order to evaluate the promoter activity of the LTR obtained from clone 6 (cl6), a test of promoter

activity using the enzyme CAT (chloramphenicol acetyl transferase) was carried out with the corresponding U3R region. In parallel, a clone containing the same U3R region of endogenous retroviral RNA expressed in normal placenta (PH74) and a clone (5M6) obtained from DNA were tested. The result presented in Figure 12 shows a very high promoter activity of the LTR derived from MS plasma (c16) and a significantly much lower activity with the sequences of non-MS endogenous origin.

EXAMPLE 2: PREPARATION OF THE C15 CLONE CONTAINING THE REGION ENCODING A PORTION OF THE MSRV-1 RETROVIRUS ENVELOPE

A RT-PCR was carried out on the total RNA extracted from virions concentrated by ultra-centrifugation of a synoviocyte culture supernatant obtained from an MS patient. The synthesis of cDNA was carried out with an oligo dT primer and the reverse transcriptase "Expand™ RT" from Boehringer according to the conditions recommended by the company. A PCR was carried out with the Expand™ Long Template PCR System (Boehringer) under the following conditions: 94°C 5 min then 93°C 1 min, 60°C 1 min, 68°C 3 min over 40 cycles and 68°C for 8 min and with a final reaction volume of 50 µl.

Primers used for the PCR:

- 5' primer, identified by SEQ ID NO: 69

5' GCC ATC AAG CCA CCC AAG AAC TCT TAA CTT 3';

- 3' primer, identified by SEQ ID NO: 116

5' TGG GGT TCC ATT TGT AAG ACC ATC TGT AGC TT 3'

A second so-called "seminested" PCR was carried out with a 5' primer situated inside the region already amplified. This second PCR was carried out under the same experimental conditions as those used for the first PCR (except that 30 cycles were used instead of 40), using 10 µl of the amplification product derived from the first PCR.

Primers used for the seminested PCR:

- 5' primer, identified by SEQ ID NO: 70

5' CCA ATA GCC AGA CCA TTA TAT ACA CTA ATT 3';

- 3' primer, identified by SEQ ID NO: 116

The primers SEQ ID NO: 69 and SEQ ID NO: 70 are specific for the pol region of MRSV-1. The primer SEQ ID NO: 116 is specific for the sequence FBd13 (also called B13) and is located in the conserved env region among the oncoretroviruses.

An amplfication product of 1932 bp was obtained and cloned in the following manner:

the amplified DNA was inserted into a plasmid with the aid of the TA Cloning kit®. The various steps were carried out in accordance with the instructions for the TA Cloning kit® (Invitrogen). At the end of the procedure, the white colonies of recombinant bacteria (white) were subcultured so as to be cultured and allow the extraction of the plasmids incorporated according to the so-called "miniprep" procedure. The plasmid preparation of each recombinant colony was cut with an appropriate restriction enzyme and analyzed on agarose gel. The plasmids possessing an insert detected under UV light after staining the gel with ethidium bromide were selected for the sequencing of the insert after hybridization with a primer complementary to the SP6 promoter present on the cloning plasmid of the TA cloning kit®. The reaction prior to the sequencing was then carried out according to the method recommended for using the sequencing kit "PRISM™ Ready Reaction AmpliTaq® FS, DyeDeoxy™ Terminator" (Applied Biosystems, ref. 402119) and the automated sequencing was carried out on the Applied Biosystems 373 A and 377 apparatus, according to the manufacturer's instructions.

The C15 clone obtained contains a region corresponding to the region of the MSRV-1 envelope of 1481 bp.

The env region of the C15 clone is represented by its nucleotide sequence (SEQ ID NO: 117) in Figure 5. The three potential reading frames of this clone are presented by their amineo [sic] acid sequence

under the nucleotide sequence. The reading frame corresponding to an MSRV-1 structural env protein is identified by SEQ ID NO: 118.

From the defined sequences obtained from clones c16 and C15, it was possible to produce a plasmid construct encoding a complete envelope followed by the 3' LTR, as presented in Figure 13 with the corresponding reading frame.

EXAMPLE 3: PREPARATION OF A 5M6 CLONE CONTAINING THE SEQUENCES OF THE 3' TERMINAL REGION OF THE ENVELOPE, FOLLOWED BY THE MSRV-1 PROVIRAL TYPE U3, R AND U5 SEQUENCES

A monodirectional PCR was carried out on the DNA extracted from immortalized B lymphocytes in culture from an MS patient. The PCR was carried out with Expand™ Long Template PCR System (Boehringer) under the following conditions: 94°C 3 min then 93°C 1 min, 60°C 1 min, 68°C 3 min over 10 cycles, then 93°C 1 min, 60°C 1 min with 15 sec of extension at each cycle, 68°C 3 min over 35 cycles and 68°C for 7 min and with a final reaction volume of 50 µl.

The primer used for the PCR identified by SEQ ID NO: 119 is 5' TCA AAA TCG AAG AGC TTT AGA CTT GCT AAC CG 3';

The primers [sic] SEQ ID NO: 119 is specific for the env region of the C15 clone.

An amplification product of 1673 bp was obtained and cloned in the following manner:

the amplified DNA was inserted into a plasmid with the aid of the TA Cloning kit®. The various steps were carried out in accordance with the instructions for the TA Cloning kit® (Invitrogen). At the end of the procedure, the white colonies of recombinant bacteria (white) were subcultured so as to be cultured and allow the extraction of the plasmids incorporated according to the so-called "miniprep" procedure. The plasmid preparation of each recombinant colony was cut with an appropriate restriction enzyme and analyzed on agarose

gel. The plasmids possessing an insert detected under UV light after staining the gel with ethidium bromide were selected for the sequencing of the insert after hybridization with a primer complementary to the T7 promoter present on the cloning plasmid of the TA cloning kit®. The reaction prior to the sequencing was then carried out according to the method recommended for using the sequencing kit "PRISM™ Ready Reaction AmpliTaq® FS, DyeDeoxy™ Terminator" (Applied Biosystems, ref. 402119) and the automated sequencing was carried out on the Applied Biosystems 373 A and 377 apparatus, according to the manufacturer's instructions.

The 5M6 clone obtained contains a region corresponding to the 3' region of the MSRV-1 envelope of 492 bp followed by the regions U3, R and U5 (837 bp) of MSRV1.

The 5M6 clone is represented by its nucleotide sequence (SEQ ID NO: 120) in Figure 5. The three potential reading frames of this clone are presented by their amineo [sic] acid sequence under the nucleotide sequence. The reading frame corresponding to the C-terminal end of the MSRV-1 env protein is identified by SEQ ID NO: 121.

EXAMPLE 4: PREPARATION OF THE LB16 CLONE CONTAINING THE REGION ENCODING THE MSRV-1 RETROVIRUS INTEGRASE

An RT-PCR was carried out on the total RNA treated with DNAseI and extracted from a choroid plexus obtained from an MS patient. The synthesis of cDNA was carried out with an oligo dT primer and the reverse transcriptase "Expand™ RT" from Boehringer according to the conditions recommended by the company. A "no RT" control was carried out in parallel on the same material. A PCR was carried out with Taq polymerase (Perkin Elmer) under the following conditions: 95°C 5 min, then 95°C 1 min, 55°C 1 min, 72°C 2 min over 35 cycles and 72°C for 8 min and with a final reaction volume of 50 µl.

Primers used for the PCR:

- 5' primer, identified by SEQ ID NO: 122

5' GGC ATT GAT AGC ACC CAT CAG 3';

- 3' primer, identified by SEQ ID NO: 123

5' CAT GTC ACC AGG GTG GAA TAG 3'

The primer SEQ ID NO: 122 is specific for the pol region of MSRV-1 and more precisely similar to the integrase region described above. The primer SEQ ID NO 123 was defined on sequences of the clones obtained during preliminary tests.

An amplification product of about 760 bp was obtained only in the test with RT and was cloned in the following manner:

the amplified DNA was inserted into a plasmid with the aid of the TA Cloning kit®. The various steps were carried out in accordance with the instructions for the TA Cloning kit® (Invitrogen). At the end of the procedure, the white colonies of recombinant bacteria (white) were subcultured so as to be cultured and allow the extraction of the plasmids incorporated according to the so-called "miniprep" procedure. The plasmid preparation of each recombinant colony was cut with an appropriate restriction enzyme and analyzed on agarose gel. The plasmids possessing an insert detected under UV light after staining the gel with ethidium bromide were selected for the sequencing of the insert after hybridization with a primer complementary to the T7 promoter present on the cloning plasmid of the TA cloning kit®. The reaction prior to the sequencing was then carried out according to the method recommended for using the sequencing kit "PRISM™ Ready Reaction AmpliTaq® FS, DyeDeoxy™ Terminator" (Applied Biosystems, ref. 402119) and the automated sequencing was carried out on the Applied Biosystems 373 A and 377 apparatus, according to the manufacturer's instructions.

The LB16 clone obtained contains the sequences corresponding to integrase. The nucleotide sequence of

this clone was identified by SEQ ID NO: 124 in Figure 11, three reading frames are determined.

EXAMPLE 5: PREPARATION OF A CLONE 2, CL2, CONTAINING IN 3' A PORTION HOMOLOGOUS TO THE POL GENE, CORRESPONDING TO THE PROTEASE GENE, AND TO THE GAG GENE (GM3) CORRESPONDING TO THE NUCLEOCAPSID, AND A NEW 5'CODING REGION, CORRESPONDING TO THE GAG GENE MORE SPECIFICALLY THE TEMPLATE AND THE CAPSID of MSRV-1.

A PCR amplification was carried out on the total RNA extracted from 100 µl of plasma from a patient suffering from MS. A water control, treated under the same conditions, was used as negative control. The synthesis of cDNA was carried out with 300 pmol of a random primer (GIBCO-BRL, France) and the reverse transcriptase "Expand RT" (BOEHRINGER MANNHEIM, France) according to the conditions recommended by the company. An amplification by PCR ("polymerase chain reaction") was carried out with the enzyme *Taq* polymerase (Perkin Elmer, France) using 10 µl of cDNA under the following conditions: 94°C 2 min, 55°C 1 min, 72°C 2 min then 94°C 1 min, 55°C 1 min, 72°C 2 min over 30 cycles and 72°C for 7 min with a final reaction volume of 50 µl.

Primers used for the PCR amplification:

- 5' primer, identified by SEQ ID NO: 126
5' CGG ACA TCC AAA GTG ATG GGA AAC G 3';
- 3' primer, identified by SEQ ID NO: 127
5' GGA CAG GAA AGT AAG ACT GAG AAG GC 3'

A second amplification by so-called "seminested" PCR was carried out with a 5' primer situated inside the region already amplified. This second PCR was carried out under the same experimental conditions as those used during the first PCR, using 10 µl of the amplification product derived from the first PCR.

Primers used for the amplification by seminested PCR:

- 5' primer, identified by SEQ ID NO: 128
5' CCT AGA ACG TAT TCT GGA GAA TTG GG 3';

- 3′ primer, identified by SEQ ID NO: 129
5′ TGG CTC TCA ATG GTC AAA CAT ACC CG 3′

The primers SEQ ID NO: [lacuna] and SEQ ID NO: [lacuna] are specific for the pol region, clone G+E+A, more specifically the E region: nucleotide position No. 423 to No. 448. The primers used in the 5′ region were defined on sequences of clones obtained during preliminary tests.

An amplification product of 1511 bp was obtained from the RNA extracted from the plasma of an MS patient. The corresponding fragment was not observed for the water control. This amplification product was cloned in the following manner.

The amplified DNA was inserted into a plasmid with the aid of the TA Cloning kit™. The 2 µl of DNA solution were mixed with 5 µl of sterile distilled water, 1 µl of a 10 times concentrated ligation buffer "10X LIGATION BUFFER", 2 µl of "pCR™ VECTOR" (25 ng/ml)and 1 µl of "T4 DNA LIGASE". This mixture was incubated overnight at 14°C. The following steps were carried out in accordance with the instructions of the TA Cloning kit® (Invitrogen). The mixture was plated after transformation of the ligation into *E. coli* INVαF′ bacteria. At the end of the procedure, the white colonies of recombinant bacteria were subcultured so as to be cultured and allow the extraction of the plasmids incorporated according to the so-called "DNA minipreparation" procedure (17). The plasmid preparation of each recombinant colony was cut with the restriction enzyme EcoRI and analyzed on agarose gel. The plasmids possessing an insert detected under UV light after staining the gel with ethidium bromide were selected for the sequencing of the insert after hybridization with a primer complementary to the T7 promoter present on the cloning plasmid of the TA cloning kit®. The reaction prior to the sequencing was then carried out according to the method recommended for using the sequencing kit "PRISM™ Ready Reaction Amplitaq® FS, DyeDeoxy™ Terminator" (Applied

Biosystems, ref. 402119) and the automated sequencing was carried out on the Applied Biosystems 373 A and 377 apparatus, according to the manufacturer's instructions.

The clone obtained, called CL2, contains a C-terminal region similar to the 5' terminal region of the clones G+E+A of MSRV-1, which makes it possible to define the C-terminal region of the gag gene and a new region corresponding to the N-terminal region of the MSRV-1 gag gene.

CL2 makes it possible to define a region of 1511 bp having an open reading frame in the N-terminal region of 1077 bp encoding 359 amino acids and a non-open reading frame of 454 bp corresponding to the C-terminal region of the MSRV-1 gag gene.

The nucleotide sequence of CL2 is identified by SEQ ID NO: 130. It is represented in Figure 6 with the potential reading frames in amineo [sic] acid.

The 1077 bp fragment of CL2 encoding 359 amino acids was amplified by PCR with the *Pwo* enzyme (5U/µl) (Boehringer Mannheim, France) using 1 µl of the DNA minipreparation of clone 2 under the following conditions: 95°C 1 min, 60°C 1 min, 72°C 2 min over 25 cycles and with a final reaction volume of 50 µl with the aid of the primers:

- 5' primer (*Bam*HI), identified by SEQ ID NO: 132
5' TGC TGG AAT TCG GGA TCC TAG AAC GTA TTC 3' (30 mer), and
- 3' primer (*Hind*III), identified by SEQ ID NO: 133
5 AGT TCT GCT CCG AAG CTT AGG CAG ACT TTT 3' (30 mer)

corresponding, respectively, to the nucleotide sequence of clone 2 at position -9 to 21 and 1066 to 1095.

The fragment obtained by PCR was linearized with *Bam*HI and *Hind*III and subcloned into the expression vectors pET28C and pET21C (NOVAGEN) linearized with *Bam*HI and *Hind*III. The sequencing of the DNA of the 1077 bp fragment of clone 2 in the two expression vectors was carried out according to the method recommended for the use of the sequencing kit

"PRISM™ Ready Reaction Amplitaq® FS, DyeDeoxy™ Terminator" (Applied Biosystems, ref. 402119) and the automated sequencing was carried out on the Applied Biosystems 373 A and 377 apparatus, according to the manufacturer's instructions.

The expression of the nucleotide sequence of the 1077 bp fragment of clone 2 by the expression vectors pET28C and pET21C are identified by SEQ ID NO: 135 and SEQ ID NO: 137, respectively.

EXAMPLE 6: EXPRESSION OF CLONE 2 IN *ESCHERICHIA COLI*

The constructs pET28c-clone 2 (1077 bp) and pET21C-clone 2 (1077 bp) synthesize, in the bacterial strain BL21 (DE3), a protein fused at the N- and C-terminus for the vector pET28C and the C-terminus for the vector pET21C with 6 Histidines, having an apparent molecular mass of about 45 kDa, identified by SDS-PAGE polyacrylamide gel electrophoresis (SDS = Sodium Dodecyl Sulfate) (Laemmli, 1970 (1)). The reactivity of the protein was demonstrated towards an anti-Histidine monoclonal antibody (DIANOVA) by the Western-blot technique (Towbin et al., 1979 (2)).

The recombinant proteins pET28c-clone 2 (1077 bp) and pET21C-clone 2 (1077 bp) were visualized by SDS-PAGE in the insoluble fraction after enzymatic digestion of the bacterial extracts with 50 µl of lysozyme (10 mg/ml) and ultrasound lysis.

The antigenic properties of the recombinant antigens pET28C-clone 2 (1077 bp) and pET21C-clone 2 (1077 bp) were tested by Westen blotting () [sic] after solubilization of the bacterial pellet with 2% SDS and 50 mM β-mercaptoethanol. After incubation with sera from patients suffering from multiple sclerosis, the sera from neurological controls and the sera from controls at the Blood Transfusion Center (CTS), the immunocomplexes were detected with the aid of an alkaline phosphatase-coupled goat serum anti-human IgG and anti-human IgM.

The results are presented in the table below.

TABLE

Reactivity of sera affected by multiple sclerosis and controls with the MSRV-1 recombinant protein gag clone 2 (1077 bp) = pET21C-clone 2 (1077 bp) and pET28C-clone 2 (1077 bp)[a]

DISEASE	NUMBER OF INDIVIDUALS TESTED	NUMBER OF POSITIVE INDIVIDUALS
MS	15	6 2(+++), 2(++), (2(+)
NEUROLOGICAL CONTROLS	2	1(+++)
HEALTHY CONTROLS (CTS)	22	1(+/-)

(a) The strips containing 1.5 µg of recombinant antigen pET-gag clone 2 (1077 bp) exhibit reactivity against sera diluted 1/100. The Western-Blot interpretation is based on the presence or absence of a specific pET-gag clone 2 (1077 bp) band on the strips. Positive and negative controls are included in each experiment.

These results show that, under the technical conditions used, about 40% of the human sera affected by multiple sclerosis which were tested react with the recombinant proteins pET28C-clone 2 (1077 bp) and pET21C-clone 2 (1077 bp). Reactivity was observed on a neurological control and it is of interest to note that the RNAs extracted from this serum, after the reverse transcriptase step, are also amplified by PCR in the pol region. This suggests that people who have not declared MS may also harbor and express this virus. On the other hand, an apparently healthy control (CTS donor) possesses anti-gag (clone 2, 1077 bp) antibodies. This is compatible with an immunity acquired against MSRV-1 independently of a declared associated autoimmune disease.

EXAMPLE 7: PREPARATION OF AN LB13 CLONE CONTAINING IN 3' A PORTION HOMOLOGOUS TO CLONE 2 CORRESPONDING TO THE GAG GENE AND IN 5' A PORTION HOMOLOGOUS TO THE 5M6 CLONE CORRESPONDING TO THE U5 LTR REGION

An RT-PCR ("reverse transcriptase-polymerase chain reaction") was carried out using total RNA extracted from virions, obtained from supernatants of B lymphocyte cells of patients suffering from multiple sclerosis, concentrated by ultracentrifugations. The synthesis of cDNA was carried out with a specific primer SEQ No. XXX and the reverse transcriptase "Expand™ RT" from BOEHRINGER MANNHEIM according to the conditions recommended by the company.

Primer used for the synthesis of the cDNA, identified by SEQ ID NO: 138:

5' CTT GGA GGG TGC ATA ACC AGG GAA T 3'

A PCR amplification was carried out with *Taq* polymerase (Perkin Elmer, France) under the following conditions: 94°C 1 min, 55°C 1 min, 72°C 2 min over 35 cycles at 72°C for 7 min and with a final reaction volume of 100 µl.

Primers used for the PCR amplification:

- 5' primer, identified by SEQ ID NO: 139

5' TGT CCG CTG TGC TCC TGA TC 3'

- 3' primer, identified by SEQ ID NO: 138

5' CTT GGA GGG TGC ATA ACC AGG GAA T 3'

A second so-called "seminested" PCR amplification was carried out with a 3' primer situated inside the region already amplified. This second amplification was carried out under the same experimental conditions as those used during the first amplification, using 10 µl of the amplification product derived from the first PCR.

Primers used for the "seminested" PCR amplification:

- 5' primer, identified by SEQ ID NO: 139

5' TGT CCG CTG TGC TCC TGA TC 3'

- 3' primer, identified by SEQ ID NO: 140

5' CTA TGT CCT TTT GGA CTG TTT GGG T 3'

The primers SEQ ID NO: 138 and SEQ ID NO: 140 are specific for the gag region, clone 2 nucleotide position No. 373-397 and No. 433-456. The primers used in the 5' region were defined on sequences of the clones obtained during preliminary tests.

An amplification product of 764 bp was obtained and cloned in the following manner:

The amplified DNA was inserted into a plasmid with the aid of the TA Cloning kit™. The 2 µl of DNA solution were mixed with 5 µl of sterile distilled water, 1 µl of a 10 times concentrated ligation buffer "10X LIGATION BUFFER", 2 µl of "pCR™ VECTOR" (25 ng/ml)and 1 µl of "T4 DNA LIGASE". This mixture was incubated overnight at 14°C. The following steps were carried out in accordance with the instructions of the TA Cloning kit® (Invitrogen). The mixture was plated after transformation of the ligation into *E. coli* INVαF' bacteria. At the end of the procedure, the white colonies of recombinant bacteria were subcultured so as to be cultured and allow the extraction of the plasmids incorporated according to the so-called "DNA minipreparation" procedure (17). The plasmid preparation of each recombinant colony was cut with the restriction enzyme *Eco*RI and analyzed on agarose gel. The plasmids possessing an insert detected under UV light after staining the gel with ethidium bromide were selected for the sequencing of the insert after hybridization with a primer complementary to the T7 promoter present on the cloning plasmid of the TA cloning kit®. The reaction prior to the sequencing was then carried out according to the method recommended for using the sequencing kit "PRISM™ Ready Reaction Amplitaq® FS, DyeDeoxy™ Terminator" (Applied Biosystems, ref. 402119) and the automated sequencing was carried out on the Applied Biosystems 373 A and 377 apparatus, according to the manufacturer's instructions.

The LB13 clone obtained contains an N-terminal region of MSRV-1 gag gene homologous to clone 2 and an

LTR corresponding to a portion of the U5 region. Between the U5 region and gag, a binding site for the transfer RNAs, the PBS "primer binding site", was identified.

The nucleotide sequence of the 764 bp fragment of the LB13 clone in the plasmid "pCR™ vector" is represented in the identifier SEQ ID NO: 141.

The binding site for the transfer RNAs, having a sequence of PBS tryptophan type, was identified at nucleotide position No. 342-359 of the LB13 clone.

As this same PBS was found in the endogenous copies homologous to MSRV1, the endogenous family thus defined is henceforth called HERV W, according to the nomenclature proposed for the endogenous retrovirus families (W=tryptophan).

A short ORF of about 65 amino acids was found in the U5 region of the 5' LTR of the LB13 clone.

Sequence of the ORF:

PMASNRAITLTAWSKIPFLGIRETKNPRSENTRLATMLEAAHHHFGSSPPLSWEL WEQGPQVTIW.

The corresponding nucleotide sequence starting at an ATG codon is capable of being expressed in a subgenomic DNA from a proviral LTR (U3RU5).

Another clone, called LA15, was obtained on the total RNA extracted from virions concentrated by ultracentrifugation from a culture supernatant of synoviocytes obtained from a patient suffering from rheumatoid arthritis. The strategy for amplifying and cloning the LA15 clone is exactly the same which was used for the LB13 clone.

The nucleotide sequence of the LA15 clone, which is represented in the identifier SEQ ID NO: 142, is very similar to the LD13 clone. This suggests that the MSVR-1 retrovirus detected in multiple sclerosis has sequences which are similar to those found in rheumatoid arthritis.

EXAMPLE 8: RECONSTRUCTION OF AN RU5-GAG REGION FROM THE CLONES LB15, LB13, CL2 AND CL17

The clones CL2 and LB13 have already been described in the preceding examples. The LB15 clone was obtained using the R sequence of the LTR of the c16 clone in order to define a primer in 5' and the anti-sense primers used are the same as for the LB13 clone. The CL17 clone was obtained by nested RT-PCR using the following primers:

5'-TCATGCAACTGCACTCTTCTGGTCCG-3' (sense)
5'-TCTTGCACTAACCTCCACTGTCCGTTGG-3' (antisense)

5'-ATCCCCCAGTAACAATTTGGTGACCACG-3' (sense)
5'-TCGGGTCTAAGAGGGTACTTCCTTTGGTAGG-3' (antisense)

The LB15 clone was obtained from virions obtained by culturing MS cells. The LB17 clone was obtained from culturing plasma from an MS patient.

These overlapping clones made it possible to reconstruct an RU5-gag sequence with a potential ORF in the gag gene, as presented in Figure 14.

EXAMPLE 9: PREPARATION OF A CLONE 87-23

The region corresponding to integrase was amplified and cloned from MS plasma using a seminested RT PCR with the following primers situated in the pol and env regions of MSRV1.

In the pol region:

5'-TTACGCAGGTCTCAGGGATGAGCTT-3' (sense-primary PCR)
5'-CGGCAGTAGCAGTCTTAGTATCTGAAGCAGTTA-3' (sense-secondary PCR)

In the env region,

5'-GGTACGGAGGGTTTCATGTAGTTTTGAG-3' (anti-sense primary and secondary PCR)

The amplified clone contains 774 bp in the pol/RT region, all the integrase region (1197 bp) and

the start of the env region (480 bp). The nucleotide sequence corresponding to the integrase region and the translation to amino acids of the potential ORF are presented in Figure 15.

EXAMPLE 10: CONFIRMATION OF THE PRESENCE OF RNA CONTAINING ENV SEQUENCES RELATED TO ERV9 IN THE RETROVIRAL PARTICLES ASSOCIATED WITH THE MSRV1 GENOME:

Sequences related to ERV9 have been found in a minor proportion in the virion preparations obtained from MS compared with the MSRV1 sequences. The existence of phenomena of co-encapsidation of phylogenetically related endogenous sequences into retroviral particles produced by a replicative strain has been described. Surprisingly, an RNA region comprising an ORF starting in the 3' portion of env and continuing potentially into the 3' LTR has been found in various MS samples. In order to specify the existence of an ORF, transcription-translation tests were carried out and made it possible to show the reality of an env ORF containing the entire transmembrane (TM) portion and ending at the start of the putative LTR. However, an additional frame (ORFX) follows and continues in the 3' LTR. The two products of expression were visualized and their respective ORFs were subcloned. Figure 16 represents the nucleotide and peptide sequences of the B13 clone already described, specifying the ORFs in the truncated env region and in the putative LTR. The presence of such RNAs may be responsible for recombinations with the replicative strain and consequently generate strains having a modified pathogenicity.

BIBLIOGRAPHIC REFERENCES

(1) Laemmli U.K. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature. (1970). **227**: 680-685.

(2) Towbin H., Staehelin T. & Gordon J. Electrophoretic transfer of proteins from polyacryalmide gels to nitrocellulose sheets: procedure and some applications. Proc. Natl. Acad. Sci. USA. (1979). **76**: 4350-4354.